

Code No: M0522/R07

Set No. 1

**IV B.Tech I Semester Supplementary Examinations, Feb/Mar 2011
DATA WAREHOUSING AND DATA MINING
(Computer Science & Engineering)**

Time: 3 hours

Max Marks: 80

**Answer any FIVE Questions
All Questions carry equal marks**

1. (a) Explain data mining as a step in the process of knowledge discovery.
(b) Differentiate operational database systems and data warehousing. [8+8]
2. Explain various data reduction techniques. [16]
3. Write the syntax for the following data mining primitives:
 - (a) The kind of knowledge to be mined.
 - (b) Measures of pattern interestingness. [16]
4. (a) What is Concept description? Explain.
(b) What are the differences between concept description in large data bases and OLAP? [8+8]
5. Explain the Apriori algorithm with example. [16]
6. Explain the following:
 - (a) Naive Bayesian classification
 - (b) Backpropagation and interpretability
 - (c) Linear and multiple regressions
 - (d) Classifier accuracy. [4+4+4+4]
7. (a) Write k-Means and k-Medoids algorithms.
(b) Explain COBWEB model. [8+8]
8. (a) Explain the classification and prediction analysis of multimedia data.
(b) What are basic measures for text retrieval? What methods are there for information retrieval?
(c) What is meant by 'authoritative' Web pages? Explain about mining the Web's link structures to identify authoritative web page. [4+6+6]

Code No: M0522/R07

Set No. 2

IV B.Tech I Semester Supplementary Examinations, Feb/Mar 2011
DATA WAREHOUSING AND DATA MINING
 (Computer Science & Engineering)

Time: 3 hours

Max Marks: 80

Answer any FIVE Questions
 All Questions carry equal marks

1. (a) Discuss about Concept hierarchy.
 (b) Briefly explain about - classification of database systems. [8+8]
2. (a) Briefly discuss the role of Data compression in data reduction process.
 (b) Briefly discuss the role of Numerosity reduction in data reduction process. [8+8]
3. Discuss the importance of establishing a standardized data mining query language. What are some of the potential benefits and challenges involved in such a task? List and explain a few of the recent proposals in this area. [16]
4. (a) What is Concept description? Explain.
 (b) What are the differences between concept description in large data bases and OLAP? [8+8]
5. Sequential patterns can be mined in methods similar to the mining of association rules. Design an efficient algorithm to mine **multilevel sequential patterns** from a transaction database. An example of such a pattern is the following "A customer who buys a PC will buy Microsoft software within three months", on which one may drill down to find a more refined version of the patterns, such as "A customer who buys a Pentium PC will buy Microsoft office within three months". [16]
6. (a) Explain the following classification methods
 - i. k-Nearest neighbor classifiers
 - ii. case-based reasoning
 - iii. rough set approach
 - iv. fuzzy set approaches.
 (b) Explain classifier accuracy. [3+3+2+2+6]
7. (a) Write k-Means and k-Medoids algorithms.
 (b) Explain COBWEB model. [8+8]
8. (a) Describe latent semantic indexing technique with an example.
 (b) Discuss about mining time-series and sequence data. [4+12]

Code No: M0522/R07

Set No. 3

IV B.Tech I Semester Supplementary Examinations, Feb/Mar 2011
DATA WAREHOUSING AND DATA MINING
(Computer Science & Engineering)

Time: 3 hours

Max Marks: 80

Answer any FIVE Questions
All Questions carry equal marks

1. Write short notes on the following:
 - (a) Efficient computation of data cube.
 - (b) Indexing OLAP data.
 - (c) Efficient processing of OLAP queries
 - (d) Metadata repository. [16]
2. Briefly discuss the role of data cube aggregation and dimension reduction in the data reduction process. [16]
3. Write the syntax for the following data mining primitives:
 - (a) Task-relevant data.
 - (b) Concept hierarchies. [16]
4. Write short notes for the following in detail:
 - (a) Measuring the central tendency
 - (b) Measuring the dispersion of data. [16]
5. Explain the following:
 - (a) Generating Association rules from frequent items.
 - (b) Improving the efficiency of Apriori.
 - (c) Constraint based Association mining. [4+4+8]
6. Explain the following:
 - (a) Scalability and decision tree induction
 - (b) Training Bayesian belief networks
 - (c) Backpropagation. [4+4+8]
7. (a) Suppose that the data mining task is to cluster the following eight points (with (x,y) representing location) into three clusters.
 A1(2,10), A2(2,5), A3(8,4), B1(5,8), B2(7,5), B3(6,4), C1(1,2), C2(4,9).
 The distance function is Euclidean distance. Suppose initially we assign A1, B1, and C1 as the center of each cluster, respectively. Use the k-means algorithm to show only

Code No: M0522/R07

Set No. 3

- i. the three cluster centers after the first round execution, and
 - ii. the final three clusters.
- (b) Explain DBSCAN algorithm with suitable example.
- (c) How does CLIQUE work? [8+4+4]
8. A heterogeneous database system consists of multiple database systems that are defined independently, but that need to exchange transform information among themselves and answer global queries. Discuss how to process a descriptive mining query in such a system using a generalization-based approach. [16]

FirstRanker

Code No: M0522/R07

Set No. 4

IV B.Tech I Semester Supplementary Examinations, Feb/Mar 2011
DATA WAREHOUSING AND DATA MINING
(Computer Science & Engineering)

Time: 3 hours

Max Marks: 80

Answer any FIVE Questions
All Questions carry equal marks

1. (a) Explain data mining as a step in the process of knowledge discovery.
 (b) Differentiate operational database systems and data warehousing. [8+8]
2. (a) Briefly discuss about data integration.
 (b) Briefly discuss about data transformation. [8+8]
3. (a) Explain the syntax for Task-relevant data specification.
 (b) Explain the syntax for Interestingness measure specification. [8+8]
4. (a) What is Concept description? Explain.
 (b) What are the differences between concept description in large data bases and OLAP? [8+8]
5. (a) Explain about constraint-based Association mining.
 (b) Give an example for Association rule mining? Classify Association rules. [8+8]
6. The following table consists of training data from an employee database. The data have been generalized. For a given row entry, count represents the number of data tuples having the values for department, status, age, and salary given in that below:

Department	status	age	salary	count
Sales	Senior	31...35	46K...50K	30
Sales	Junior	26...30	26K...30K	40
Sales	Junior	31...35	31K...35K	40
Systems	Junior	21...25	46K...50K	20
Systems	Senior	31...35	66K...70K	5
Systems	Junior	26...30	46K...50K	3
Systems	Senior	41...45	66K...70K	3
Marketing	Senior	36...40	46K...50K	10
Marketing	Junior	31...35	41K...45K	4
Secretary	Senior	46...50	36K...40K	4
Secretary	Junior	26...30	26K...30K	6

Let salary be the class label attribute.

Design a multilayer feed-forward neural network for the given data. Label the nodes in the input and output layers. [16]

7. (a) Suppose that the data mining task is to cluster the following eight points (with (x,y) representing location) into three clusters.

Code No: M0522/R07

Set No. 4

A1(2,10), A2(2,5), A3(8,4), B1(5,8), B2(7,5), B3(6,4), C1(1,2), C2(4,9).

The distance function is Euclidean distance. Suppose initially we assign A1, B1, and C1 as the center of each cluster, respectively. Use the k-means algorithm to show only

- i. the three cluster centers after the first round execution, and
 - ii. the final three clusters.
- (b) Explain DBSCAN algorithm with suitable example.
- (c) How does CLIQUE work? [8+4+4]
8. (a) Explain spatial association analysis.
- (b) How can we study time-series data? Explain.
- (c) What is sequential pattern mining? Explain.
- (d) Explain text data analysis and information retrieval. [4+4+4+4]
